DOCUMENT RESUME

ED 230 622 TM 830 469

AUTHOR Torres, Rosalie T.; Harnisch, Delwyn L.
TITLE Functional Literacy Testing in the United States: Measurement Issues and Theoretical and Practical Implications.
PUB DATE Apr 83
NOTE 21p.; Paper presented at the Annual Meeting of the American Educational Research Association (67th, Montreal, Quebec, April 11-15, 1983).
PUB TYPE Speeches/Conference Papers (150) -- Reports - Evaluative/Feasibility (142) -- Information Analyses (070)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Adults; Criterion Referenced Tests; Educational Diagnosis; Elementary Secondary Education; *Functional Literacy; *Measurement Techniques; Minimum Competencies; *Minimum Competency Testing; Remedial Instruction; *Test Theory; *Test Use; *Test Validity
IDENTIFIERS *Content Validity

ABSTRACT

A review of functional literacy testing in the United States from 1955-82 is provided by summarizing the results of literacy assessment studies and synthesizing the major issues which they have engendered. These reviews are grouped with respect to content, criterion-related, and construct validity. The paper concludes by (1) summarizing some of the theoretical and practical implications of these findings and (2) citing measurement techniques which provide information that can enhance the use of existing functional literacy tests. (Author/PN)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED230622

Functional Literacy Testing in the United States:

Measurement Issues and Theoretical and Practical Implications

Rosalie T. Torres and Delwyn L. Harnisch
University of Illinois, Urbana-Champaign

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☒ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

R. T. Torres

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

Paper presented at the annual meeting of the American Educational Research Association, Montreal, April 1983.

"Printed in U.S.A."

TM 830 469

Abstract

This paper examines the literacy problem in the United States in terms of the measurement issues which have been raised about functional literacy testing. After a brief statement of the literacy problem the major efforts addressing it are discussed. A review of the measurement critiques of these studies is then presented. These critiques are grouped with respect to content, criterion-related, and construct validity. The paper concludes by (1) summarizing some of the theoretical and practical implications of these findings and (2) citing measurement techniques which provide information that can enhance the use of existing functional literacy tests.

Functional Literacy Testing in the United States:

Measurement Issues and Theoretical and Practical Implications

In this century, widespread concern about the literacy levels of U.S. adults typically has followed some event calling for an assessment of literacy levels for a sizeable portion of the population. The military mobilizations of WW-I and WW-II requiring the testing of recruits for induction into service have been such events. The results from these assessments led to an awareness of and concern over the apparently low literacy levels of these recruits, and thus, their ability to perform basic military jobs (Hunter & Harman, 1979). More recently the switch to a volunteer army has created the same concern about the literacy levels of volunteers.

The rather discouraging findings from these specific literacy assessments as well as the negative trends seen in ongoing assessments of literacy (e.g., high school students' performance on standardized achievement tests) have all contributed to the currently raging concern over basic reading, writing, and mathematics skills of adults in the U.S. In addition, the educational system's newest area of responsibility and concern is adults' attainment of basic knowledge and skills in the use of computers for coping with the information age (Aron, 1982).

These kinds of concerns imply that our society's expectations about what literacy levels in the U.S. ought to be are not being met. An underlying assumption is that more literacy is good. Social policy makers have long associated the attainment of literacy with enhanced individual and societal development (A Bibliographic Guide to Functional Literacy, 1979; Hunter & Harman, 1979). This assumption, however, has been held implicity, and therefore subjected to little if any critical examination by decision makers.

Nonetheless, spurred on by this biting awareness of the disparity between observed and expected levels of literacy, and the notion that increasing literacy is a desirable goal, beginning in the late 1960's and early 1970's major efforts were undertaken to do something about the literacy problem in the U.S. Broadly speaking, these efforts were attempts to define and measure the quality and extent of literacy in the U.S.

The purpose of this paper is to provide a review of functional literacy testing in the U.S. by summarizing the results of these studies and synthesizing the major issures which they have engendered. This review addresses validity issuses in the measurement of literacy and provides a basis from which important theoretical and practical implications are considered. The paper concludes by discussing recent developments in the use of existing functional literacy tests.

MAJOR EFFORTS ADDRESSING THE LITERACY PROBLEM

The early considerations about measuring literacy were tied to the simplest meaning of literacy, that is, reading and writing ability. In keeping with this notion, the criteria for literacy consisted of reading

and writing abilities equivalent to skill levels of the upper elementary grades. Departure from this definition was evidenced in the initial plans for the National Right-to-Read Project of the early 1970's. The intent of this project was to "ensure that every person arriving at adulthood will be able to read and understand the whole spectrum of materials that one is likely to encounter in daily life" (Olson, 1975). Although the project was never completely realized, its conceptualization resulted in assessments of adult literacy being made independently of grade level equivalents (Peck & Cling, 1977).

The efforts to define and measure literacy that followed the Right-to-Read Project all emphasized the concern that individuals can not only read and write, but also that they can use these skills in integrating information necessary for independent survival or functioning in our society. This notion of literacy skills for survival or coping is embodied in the now popular term, functional literacy. More specifically, it has been defined as relating to levels of skill that individuals or populations need in order to complete some specified real-life reading task (Kirsch & Guthrie, 1977). Before describing the studies which addressed literacy in this way it is important to note that they are part of two other recent movements in educational assessment--minimum competency and criterion referenced testing.

First, measures of functional literacy are concerned with assessing skills that are adequate for functioning, that is, minimum competencies. Examinees must demonstrate minimum proficiencies with respect to some specified literacy skills or standards of performance. Second, minimum competencies are usually assessed by criterion-referenced tests. This follows from the fact that test items can be related to specific literacy skills or objectives and thus, the test results describe an individual's performance with respect to these objectives or behaviors. Thus, functional literacy tests provide a means of assessing minimum levels of literacy skill and describe individuals' performance with respect to these skills.

Harris and Associates Studies

The Survival Literacy Study (1970) and the 1971 National Reading Index: A Study of Functional Reading Ability in the U.S. (1971) were conducted by Louis Harris and Associates. Both sides attempted to define literacy in terms of the ability to respond to practical tasks of daily life. The first study attempted to assess the abilities of U.S. adults to read and fill out common application forms. Test results indicated that from 4.3 million to 18.5 million Americans are functionally illiterate. Performance on these practical abilities increased in direct proportion to income and years of education, and decreased in proportion to age. Women and people in the West exhibited the fewest reading deficiencies.

The second study attempted to assess the abilities of U.S. adults to use common printed materials such as telephone dialing instructions and newspaper advertisements. Interview results indicated that 4% of the U.S. adult population suffers from serious reading deficiencies. Again, performance on these abilities increased in proportion to income and

education. Whites and people in the West and Midwest exhibited the fewest deficiencies.

Adult Functional Reading Study

The Adult Functional Reading Study, conducted by the Educational Testing Service was concerned with describing typical adult reading tasks (Murphy, 1973). It defined reading competency as those skills suitable for adequate functioning in normal day-to-day living. Some of the results were analyzed by Kirsch and Guthrie (1977) and indicated that one out of five respondents of a national survey could not complete maintenance reading tasks (tables of contents, common signs, and train schedules, etc.); while one out of four respondents could not complete occupational tasks (employment applications, sick leave requests, etc.). Respondents' reading habits were differentially associated with socioeconomic status.

Project Realistic

Project Realistic related reading, listening, and arithmetic skills to indices of job proficiency (Sticht, Caylor, Kern, & Fox, 1972). It was designed to determine functional literacy levels for selected military jobs having civilian counterparts. The study used empirical methods for establishing levels of general reading ability needed for successful job training and performance. As a result, a comparison between these reading requirements and stated job requirements could be made, and training programs could be designed to better prepare students for reading skills needed in various jobs (Sticht & McFann, 1975).

Reading/Everyday Activities in Real Life

Reading/Everyday Activities in Real Life (R/EAL), a competency based test of functional literacy was developed by Lichtman (1974). It samples from a domain of real life reading tasks and is designed to provide diagnostic and prescriptive information for examinees aged 10 years and older. The reading tasks assessed included sets of directions, maps, application forms, and technical documents.

Mini-Assessment of Functional Literacy

The Mini-Assessment of Functional Literacy (MAFL) reported by the National Assessment of Educational Progress (1974) emphasized reading materials frequently encountered in everyday life. It attempted to assess the functional literacy of U.S. 17-year-old high school students using materials with which individuals must cope in order to function adequately (passages, pictures, labels, maps, graphs, forms, and reference materials). The results of examinees grouped by region, sex, race, parental education, and size and type of community indicated that all groups performed best on understanding word meanings and worst on drawing inferences (Kirsch & Guthrie, 1977).

Adult Performance Level Study

The Adult Performance Level (APL) study conducted by the University of Texas in the mid-1970's defined functional competency, a broader term

than functional literacy. This study was the most comprehensive in terms of its goals and impact. It defined functional competency as coping behaviors consisting of the application of a set of skills to a set of general knowledge areas (Final Report: The APL Study, 1977). The skills were communication, computation, problem solving, and interpersonal relations; while the knowledge areas related to occupations, consumer economics, community resources, government and law, and health. Thus, functional competency was seen as incorporating skills beyond those needed for dealing with the printed materials encountered in daily life.

In addition to developing this definition, the study attempted to (1) assess the functional competence of adults in the U.S., (2) develop a device for providing diagnostic and prescriptive information for examinees, and (3) initiate activities toward developing adult education curricula which teach the functional competencies defined. While traditional achievement tests (e.g., the General Educational Development Test) measure the reading, writing, and computational skills taught in school, a major intent of the APL Study was to provide a more direct means of assessing survival skills.

The study developed a scale of three levels of competency which are associated with different levels of adult success. Success was defined in terms of a conjoint index of income, job status, and education. The results of a national survey using the APL Test estimated (1) 19.1% of the U.S. adult population to be functionally incompetent; (2) 33.9% to be marginally competent; and (3) 46.3% to be highly competent. Females, whites, and persons living in the Northeastern, Northcentral, and Western parts of the U.S. performed the best on the APL Test.

Competency Testing Programs

High school competency testing programs have been established by a majority of state departments of education. These programs have been introduced as a means of bolstering educational standards and remedying the problems of social promotion and grade inflation, all of which are perceived to contribute to the incidence of illiteracy in the U.S. (Haney, 1983; Madaus, 1983). More specifically, their objective is to assess students' attainment of minimum proficiency in skills necessary to function as adult citizens. The states have developed tests for these proficiencies which address the traditional basic skills of reading, writing, and mathematics; and in some states high school students must perform above specified cut-off scores in order to receive diplomas (Rogers, 1981). For example, in Florida, following three administrations of the State Assessment Test-II (previously called the Florida Functional Literacy Test) to the graduating class of 1979, 1.9% of the white seniors and 20.0% of the black seniors had not received scores above the cut-off point established for receipt of a standard high school diploma (Pullin, 1983).

Purposes of these Efforts

Any one of these efforts to study functional literacy may have had one or more of three purposes, either implicitly or explicitly stated. These purposes are (1) description, (2) certification, and (3) diagnosis

and remediation; and are somewhat loosely tied to the goal of increasing functional literacy in the U.S.

Description

The most obvious purpose is description. It was referred to earlier as the impetus behind most of the major attempts to study functional literacy: to describe the levels of functional literacy for either all U.S. adults or for some subgroups of them. This purpose satisfies the society's need for normative information about functional literacy by providing descriptive information from which social political, and ultimately, funding decisions are often made.

In order to describe the levels of functional literacy, functional literacy itself must first be described. Thus, the descriptive purpose consists of both a definitional and a measurement component. The definitional component is a description of what is meant by functional literacy. The measurement component is a description of the levels of functional literacy as defined.

In as far as each of the projects mentioned above included descriptions of functional literacy, they all served the definitional component of this purpose. In also providing estimates of literacy levels in the U.S., the Harris and Associates Studies, the Adult Functional Reading Study, the MAFL, the APL Study, and the competency testing programs served the measurement component of this purpose as well.

Certification

A second purpose is certification, that is, to determine whether a person is functionally literate depending upon their performance on a test of functional literacy. This is in part the intent behind some states' high school graduation requirements which stipulate that seniors pass functional literacy tests. Many tests of functional literacy can be used for this purpose by determining a cut off score above which examinees are deemed functionally literate. The R/EAL and APL Tests, as well as tests being developed by individual states' competency testing programs, are being used for this purpose.

Diagnosis and Remediation

A third purpose is diagnosis and remediation. It consists of ascertaining a description of a person's literacy skills from their performance on a test and then using that description to provide corrective instruction. Thus, through an examination of test items missed and knowledge about how each item relates to specific functional literacy skills, remediative instruction is designed for examinees.

Tests of functional literacy can be used for diagnosis and remediation in preparing both high school and adult education students for various certification requirements. Almost any test of functional literacy can be used at least to some extent for this purpose provided

that information about the relationship between test items and specific functional literacy skills is available.

MEASUREMENT ISSUES

In summary, we have witnessed several major efforts addressing the literacy problem in the U.S. They defined literacy in a variety of ways and attempted to serve a number of purposes which are loosely tied to an implicit goal of increasing functional literacy in the U.S. Not surprisingly, the development of critical issues regarding the measurement of literacy in the U.S. has closely followed these attacks on the problem. In turn, the measurement issues have led to larger concerns about the definition of and the goals for literacy in the U.S.

The immediate measurement issues have arisen out of concerns about the results of these studies and their realized or potential purposes, both of which were just described. Studies such as the Harris and Associates studies and the APL study implied that sizeable portions of the U.S. adult population do not possess adequate coping skills for daily living in our society. These results are typically differentially associated with low education, low income, minorities, males and the Southern U.S. The purposes for which these findings can be used have serious implications for individuals, especially when decisions about certification and/or appropriate remedial instruction are being made. As a result, a number of researchers have examined these studies with respect to several of the classical measurement concerns in an attempt to evaluate the meaning and accuracy of the studies' findings.

These critiques are reviewed in the next section. The discussion is divided into three areas of measurement concern: content, criterion-related, and construct validity. This classification has been made as a matter of convenience. As will be shown, and recently argued by persons in the measurement community (see Madaus, 1983) validity concerns are inextricably related. Due to its broad scope and level of government support, most of the measurement critiques address the APL Study. However, the issues addressed in this study are representative of the measurement concerns for most all of the studies.

Content Validity

Establishing content validity is generally a first concern in the development of all tests (Allen & Yen, 1979). Concern with the analysis of the content of a test is logically dealt with at the time that test items are written. Classical test writing procedures referred to by Bormuth (1975) involve the definition of a content domain and subsequent sampling from it. The types of items written are sampled from a domain defined by a matrix of subject areas-by-mental processes. The mental processes are those used to exhibit knowledge of the subject areas.

This procedure suggests several areas of investigation for determining the content validity of criterion-referenced tests. Some of these areas are (1) the clarity of the domain definition; (2) the sampling adequacy of the test content; and (3) the technical soundness of the test items (Fitzpatrick, 1981). These considerations are important

for several types of content validity which are discussed in detail below.

Results of Content Validity Studies

Face validity. Face validity is the simpler form of content validity involving subjective agreement among persons that a test measures the relevant trait(s). Since functional literacy tests are criterion-referenced competency-based assessments of skills necessary for everyday living, it is reasonable to expect that they might possess content validity, especially face validity. In their review of the APL Test, Haney and Lloyd (1980) conclude that one of the major contributions of the test is its attractiveness to participants in adult basic education classes. From an examination of the test and curriculum materials related to it, participants are readily able to see the relevance of these materials to concerns in their daily lives, especially in comparison to more traditional tests like the GED. Likewise, in their critique of the APL program, Griffith and Cervero (1977) conclude that the versions of the APL Test do possess face validity.

Sampling validity. However, functional literacy tests are more likely to be investigated regarding logical or sampling validity, the more vigorous form of content validity. It involves the logical design of items to cover all the important areas of the domain defined (Allen & Yen, 1979). A few reviewers have judged functional literacy tests favorably on the types of criteria listed above. Heller (1978) indicates that the APL Test possesses some degree of content validity based upon its skill-by-content area definition of the domain of functional competency. In their review of functional literacy, Nafziger, Thompson, Hiscox, and Owen (1975) rate the APL Test as fair with respect to validity. This rating is in part based upon the sampling of the behavior domain, the technical review of the test by experts, and several cycles of field testing and redesigning. Anders (1981) as well as Nafziger et al. rate the R/EAL test favorably with respect to content validity based upon the specification of the relationship between test items and test objectives.

Tests of functional literacy do not always fare well under scrutiny for sampling validity. Griffith and Cervero (1977) maintain that the APL Test developers provide no evidence for the sampling validity of the test, i.e., that its items represent an assessment of five content and five skill areas.

Domain definition. Other reviewers of the APL Test (Heller, 1978; Kirsch & Guthrie, 1977) question how functional competency is defined. The major issue is whether it is defined in terms of success or survival; each of which is seen as a different criteria with different implications for test validity as well as for pedagogical and philosophical considerations. Heller points out that if the definition is in terms of survival, then content validity is important in establishing what all adults should know in order to survive. The interpretation of a test based upon a survival definition is that persons not passing it lack some essential skills. Thus, an important content validity consideration would be the quality of the test items. Along these lines, Heller

reports that 39 of 64 components of the APL Test remained unchanged after a national assessment of competency levels using the test.

Both the adequacy of the definition of functional literacy and the sampling of items from the definition have been questioned for the R/EAL Test; the tests used in the Adult Functional Reading Study, the Survival Literacy Study, the MAFL; and Project Realistic (Caughran & Lindolf, 1972; Kirsch & Guthrie, 1977).

Linking tests to instruction. A more recent criterion for the determination of content validity stems from the certification (and prediction) use of functional literacy tests in high schools. This criteria requires that the items and objectives of a functional literacy test match the instruction provided to students. In the class action suit brough against Florida's use of a minimum competency test for high school graduation, measurement experts testifying for the plaintiffs argued that content validity should address the question of whether the test specifications, objectives, and items cover material in the school curriculum that students have been taught (Haney, 1983).

Anderson and Anderson (1981) attribute the poor performance of high school students on a functional literacy test in mathematics to the test's lack of content validity. The curriculum and instructional emphases were found to be unrelated to the goal areas of the test.

McClung (1977) calls this type of content validity instructional validity. In discussing the possible legal challenges against he APL Test and minimum competency tests in general, he warns that these tests must be evaluated in this way before they are used as a basis for promotion and graduation in secondary schools. The appeals court for the Florida case supported this position by agreeing with the plaintiffs that the issue of the content validity of the Florida test encompassed the question of whether the test covered what was taught in Florida schools (Haney, 1983).

Discussion of Content Validity Studies

These varied but generally negative findings regarding the content validity of functional literacy tests stem from the inherent difficulty in defining functional literacy. This definitional problem exacerbates the effects of (1) the subjectivity of content validity judgments and (2) the variety of purposes for which functional literacy tests are being used. Since content validity judgments are made largely via subjective analyses, they will be as varied as there are different opinions about the definition of functional literacy. Cervero (1980) maintains that the APL Test does not possess content validity, not so much because the test is technically unsound, but because it is not logically possible to define a universe of behaviors for functional competency without respect to a value position. And relatedly, while the appeals court upheld the lower court decision that the Florida graduation test had construct validity (on the basis of its adequate measurement of functional literacy as defined by the State Board of Education), it did recognize that functional literacy had not been defined in a way acceptable to to either all educational academicians or the public (Haney, 1983).

The problem of defining functional literacy is related to content validity and the purpose of functional literacy tests in at least two ways. First, as previously indicated, content validity is most important if the definition of functional literacy is in terms of survival. In this case, the purpose of the test is description and validation studies should concentrate on the integrity of test items. However, as Heller (1978) further delineates, if functional literacy is defined in terms of success, then the research concern is with identifying what successful adults know that nonsuccessful adults do not know. The purpose of the test is prediction, and validation studies should concentrate on the discriminatory power of the test.

Second, the concern with the instructional content validity of functional literacy tests is related to the specificity of definitions of functional literacy. This specificity is consistent with underlying raison d'etre of functional literacy tests: assessing specific skills for daily living. However, it is because of this specificity that functional literacy tests are being challenged to show instructional content validity when they are used as high school graduation requirements which assume that the tests are predictive of success or survival in adult life. Yet, this specificity mitigates the predictive validity of the tests for students who happen not to have received the relevant instruction but have the ability to learn how to function in life when the need arises (Anderson & Anderson, 1981). As will be shown in the following section, these definitional and related problems manifest themselves in considerations of other types of validity as well.

Criterion-Related Validity

Criterion-related validity is especially important for functional literacy tests since they are often used for making decisions about individuals with respect to certification and/or remediative instruction. The rationale for making a decision at all is based upon the validity of the relationship of the test to other criterion or behaviors which the test is purported to predict. Among the areas of investigation for criterion-related validity are (1) the correlation between the test scores and the independent criterion measure which the test is attempting to predict, (2) the measurement of the test itself or the criterion or both, and (3) the determination of cut-off scores used to decide whether or not an examinee possesses the trait being measured or the criterion being predicted.

However, much of the concern with the validation of functional literacy tests has focused on the content rather than on the concurrent or predictive validity of the instruments. Amongst the validity challenges brought against the Florida graduation test, little attention was given to the relevance of criterion-related validity in considering the fairness of the test (Haney, 1983). Generally speaking, an emphasis on the definition of functional literacy is perhaps reasonable to expect considering the time and financial resources required for collecting the necessary data on related variables for the latter type of validation. Nonetheless, Linn (1979) argues that evidence for content validity is not sufficient for criterion-referenced tests; particularly because scores on the tests are likely to be used as a basis for differential treatment of individuals.

Results of Criterion-Related Validity Studies

For the most part, tests of functional literacy have not been reviewed favorably regarding criterion-related validity. As previously mentioned Nafziger et al. (1975) did rate both the APL Test and the R/EAL Test fair on validity. However, their assessment was based primarily on concerns of content and not criterion-related validity. That part of their evaluation which did address criterion-related validity included (1) mention of the research showing the relationship of APL Test items to criteria such as education, job status, and income and (2) mention of the correlation between the R/EAL Test and the Stanford Achievement Test.

While the validity coefficient of .74 between the R/EAL Test and the Stanford Achievement Test has been referred to favorably by at least one researcher (Anders, 1981); the validity coefficients of .56 (years of schooling), .33 (occupational status), and .39 (family income) for the APL Tests have been questioned in a number of ways. Griffith and Cervero (1977) and Cervero (1980) maintain that the APL Test lacks concurrent validity because (1) these correlations are insubstantial and (2) education, occupation, and income lack conceptual clarity as a definition of success. Concern about the specification of success as the criterion for the APL Test has been voiced by others as well. Kirsch and Guthrie (1977) criticize this criterion as unrepresentative of the range of behaviors specified in the APL content-by-skill taxonomy.

Relatedly, Heller (1978) charges that correlations with education, occupation, and income do not show the test to be related to success, particularly since substantial numbers of successful persons answered a number of the items incorrectly. Moreover, any test which assesses reading, writing, and computational skills would correlate highly with education, occupation, and income.

Finally, Haney and Lloyd (1980) conclude that overall the APL Test items have higher correlations with ethnicity, age, and city size than they do with education, occupation, and income. They report that the percentages of items which correlate significantly with education, occupation, and income are 100%, 72%, and 43% respectively; while the percentages of items which correlate significantly with ethnicity, age, and city size are 95%, 86%, and 99% respectively. These relationships were not addressed by the test developers.

The determination of cut-off scores for functional literacy tests has also been addressed in the review literature, and not surprisingly by now, unfavorably. The literacy or competency levels of both the APL and the Survival Literacy Studies have been deemed uninterpretable because it is not clear how the levels relate to the criteria selected to distinguish among them (Kirsch & Guthrie, 1977). It is important, however, to analyze how these cut offs were determined because they result in the establishment of labels for individuals upon the basis of which decisions are made. In this regard, some researchers (Bunch, 1978; Fitzpatrick, 1981) are advocating methods for determining cut-off scores which take into account the relative severity of making either false positive or false negative errors in classificatory decisions.

Discussion of Criterion-Related Validity Studies

The investigation of criterion-related validity of functional literacy assessments concentrates primarily on the measurement of the criterion. Although validity coefficients have been considered, the real concerns relate to the construct validity of the criteria. The APL Test developers maintain that functional competence is directly related to success in adult life; yet, it is not at all clear that measuring education, occupational status, and family income is sufficient as a measure of success in our society. Equating these variables with successful and nonsuccessful persons results in persons with less education, lower job status, and lower income being classified as unsuccessful at the lower APL's. Thus, the test fails to discriminate between incompetence and lack of education, job status, or income (Haney & Lloyd, 1980). These problems in the measurement of the criteria, and therefore in the relationship between the test and the criteria make any use of cut-scores based on this relationship questionable.

Moreover, the finding that large numbers of successful persons answer some items incorrectly has called into question the meaning of the APL's. Unexpected results such as this have led to an examination of these items and the skills they are measuring, as well as further investigations of the construct validity of functional literacy tests. In citing criticisms of APL Test items which suggest how difficult it is to find commonly encountered life-skills problems, Kronick (1981) concludes that it is dangerous to lose sight of the actual test questions when interpreting test results and formulating conclusions.

Construct Validity

Establishing the construct validity of criterion referenced tests is particularly important because the use of criterion referenced tests involves making inferences about individual's skills, knowledge, abilities, or behaviors on the basis of their performance on the test (Fitzpatrick, 1981). If the test does not measure the skills or traits that it purports to measure then these inferences are not accurate.

Among the ways to assess construct validity are (1) examination of group differences on the test, (2) factor analysis, (3) studies of change in conditions relevant to the underlying trait, (4) studies of the psychological processes involved in individuals' test performance, and (5) studies or correlations among tests and/or items assumed to measure the same or different traits (Cronbach & Meehl, 1955). Thus, evidence for construct validity is provided by a variety of methods often used to establish other types of validity.

Results of Construct Validity Studies

Information provided by the test developers suggest that the constructs behind the original APL Test were that skills and knowledge essential for success can be measured on a two-dimensional model which assess both skills and content knowledge (McClung, 1977). Cervero (1980) specifically investigated the construct validity of the APL Test using factor analysis. His intent was to determine whether the test items measure the skill areas they were designed to measure, and thus, whether

the test measures all skill areas that it purports to measure. The results indicate that the test measures three areas: verbal ability, writing, and computation. Problem solving does not appear as a separate skill; rather, examinees probably use the three skills identified to answer items keyed to problem solving. In addition, it was found that measurement of these skills was invariant across the five content areas. Thus, performance on the content area items explains no more variance than does performance on the three skill areas.

These results were further confirmed by another study in which Cervero (1981) compared the GED and APL Tests. A high positive relationship between the two sets was shown by a correlation of .81. In addition, factor analyses performed on this sample of APL Test data indicated that one trait, reading ability, accounts for performance on the test. Previous research has identified this same factor as critical to performance on the GED Test. Cervero concludes that the skills for competent functioning as defined by the APL Test are the same as those measured by traditional tests such as the GED, and thus, constitute the long term objectives of high school education.

As referred to earlier, further investigations of construct validity were made with the examination of items on which persons performed poorly. The intent was to examine the cognitive processes invoked in responding to these items. Many of the items required factual information for the correct response. This brings up the question of whether possessing specific information is the best measure of coping skills in our society. As Heller (1978) points out, a better indication of competency is more oriented toward process skills, that is, knowing how to use resources which can provide answers.

Finally, as mentioned in the discussion of content validity, the appeals court in the Florida case found that the Florida test did have adequate construct validity. This determiniation was made on the basis of the test's measurement of functional literacy as defined by the State Board of Education (Haney, 1983).

Discussion of Construct Validity Studies

Overall, the construct validity assessment of functional literacy tests have consisted of factor analytic studies indicating that the familiar process skills of reading, writing, and arithmetic are being measured. The resulting implication that process type skills may be most important for competent functioning, is also supported by the findings that (1) large numbers of persons perform poorly on highly specific items and (2) that further variance in test performance is not explained by knowing the content areas for which items were written. Persons possessing more generalizable skills will better be able to cope with the practical situation in their daily lives which are rapidly changing with respect to content (Cervero, 1981).

This issue of generalizable vs. specific skills is behind the concern for the instructional content validity of functional literacy tests. If students are not taught the specific information which some of the test items require they will not be able to perform well regardless

of the fact that test performance also requires reading comprehension skills.

THEORETICAL IMPLICATIONS OF MEASUREMENT ISSUES

Validation studies of almost any type provide information of theoretical importance for the traits being examined. This is particularly the case with the validation studies of functional literacy tests. Investigations of content and criterion-related, as well as construct validity, have led to conclusions which more than anything else point to questions about the theoretical underpinnings of functional literacy. These conclusions and questions are summarized in the following sections.

Definition of Functional Literacy

Examinations of the definitions of functional literacy which are adhered to by different functional literacy assessment devices have made it clear that there is no one way to define all of the behaviors implied by functional literacy for any one group. Any definition must be made within the contexts of specific groups, their goals, and the skills required to meet these goals.

Defining functional literacy includes making statements regarding what is important about functional literacy in our society, that is, what are the observable behaviors that should relate to it? Do they have to do with success or survival? And if so, how does success or survival relate to functional literacy for different groups? The questions linger: what is literacy and how much of it do you have to have before you are considered functional (Ayrer, 1977)? Clearly, the answers are dependent upon the respondent's values, expectations, and standards.

For these reasons decisions to use a functional literacy test with a particular group must consider the appropriateness of its definition of functional literacy for that group. The way to view any test at all is to consider the hidden value system built into the test (Ayrer, 1977). In this regard Kirsch and Guthrie (1977) suggest that better instruments to measure literacy be designed and that they be tied to specific purposes and decisions which expect to be made on the basis of their results. For example, these authors suggest that, highly specific tests might be developed for (1) determining minimal skills of high school graduates or (2) providing employers with information from personnel selection and placement. Validation procedures then could concentrate on the appropriateness of the test for specific purposes.

Measurement of Functional Literacy

Studies of functional literacy have been concerned that persons are competent in performing skills adequate for daily living. Willingham (1980) maintains that assessing competence is assessing the ability to carry out a task successfully rather than to accumulate discrete bits of knowledge. This extension of knowledge that competence implies requires divergent and intergrating skills. From the validity investigations reviewed in this paper it can be seen that these skills are not easy to measure with conventional tests. This fact has certainly been borne out

by the finding that at least one of the functional literacy/competency assessments measure skills no different than those which conventional education tests measure.

Thus, functional literacy tests differ from traditional tests only in the content of their items. The practical nature of this content has, in fact, been cited as the major contribution of tests such as the APL Test. Indeed, students have been observed to enjoy the tests because of their greater appeal in this regard. And although this contribution may be important for motivating individuals to master the curriculum tied to the objectives of the test, it has not redeemed functional literacy testing. The educational community's most substantial challenge is determining whether any standard of validity is high enough that educators can support the use of a single test to make critical decisions about individuals (Pullin, 1983). As the National Academy of Education's (1978) Committee on Testing and Basic Skills concluded, "The effort to determine and ensure minimum competency standards for high school graduation will...fail of its own weight, for the scaffolding of existing test designs is too weak to carry such an emotionallky-laden and ambiguous burden" (p. 9; cited by Madaus, 1983).

In light of these realizations a number of alternative means for addressing the literacy problem in the U.S. are being proposed. Among these are maintaining profiles of students' qualities, long-term, and short-term goals, and discussing these profiles with students throughout their education. The profiles would help to alert students to their poorly developed skills as well as to help develop programs to meet students' needs (Anders, 1981).

USE OF EXISTING FUNCTIONAL LITERACY TESTS

Serious questions have been raised about the validity of functional literacy tests. The answers to these questions have challenged the accuracy of the test results and indicated ways in which the tests may be biased against and inappropriate for different groups of individuals. As a result, professionals in the field are suggesting (1) that further efforts be made to define functional literacy, (2) that new tests for specific pruposes and groups be developed, and (3) that alternative means be developed for addressing the literacy problem. However, responding to these suggestions and implementing procedures based upon them will take time and effort. Meanwile, functional literacy tests such as those described in this paper are being used.

However, several efforts to make better use of the existing tests have recently been reported. These efforts involve the examination of errors made by respondents. First, Harnisch (1982) used APL Test data to show how unusual response patterns on groups of test items may render the standard interpretation of the test score misleading. Unusual response patterns were represented by a numerical index which reflects the degree to which examinees answer very easy items incorrectly but answer some very difficult items correctly. There are a variety of factors which may account for such response patterns. They include guessing, test anxiety, an unusual instructional history or experiential background, or copying another examinee's answers to particular items. By examining an individual response pattern and developing explanations for any

unusualness in it, it is possible to know whether usage of that person's test score is fair or appropriate. Looking at response patterns for groups can help to indicate when instruction has not been relevant to the objectives of the test.

Second, Fisher (1981) used test data collected from the functional literacy test developed in the Adult Functional Reading Study to show how certain categories of errors are consistent with errors in information processing. He posits that information processing errors are made when large numbers of functionally literate persons answer items incorrectly. These types of errors are due to carelessness, and not deficits in knowledge or skills. Knowing which type of error an examinee has made provides information about the most appropriate type of remedial instruction. Knowledge deficit errors can best be remediated by having students practice with items similar to those missed. Processing errors can best be remediated by having students reread items identical to those missed.

In summary, this paper has attempted to review a number of the measurement issues which are being addressed in functional literacy testing. One area of concern deals with the validation of functional literacy tests. The examination of validity issues has helped to point out several directions for further theoretical development. The other area of concern deals with identifying the nature of errors made on functional literacy tests. These efforts have shown how information provided by analyses of errors can enhance usage of existing functional literacy tests.

References

Allen, M. J., & Yen, W. Introduction to measurement theory. Monterey, CA: Brooks/Cole, 1979.

Anders, P. L. Test review: Tests of functional literacy. Journal of Reading, 1981, 24(7), 612-619.

Anderson, L. W., & Anderson, J. C. Functional literacy tests: A case of anticipatory validity? Educational Evaluation and Policy Analysis, 1981, 3(2), 51-55.

Aron, H. The impact of computers on literacy. Lifelong Learning: The Adult Years, 1982, 6(4), 8-9, 22.

Ayrer, J. E. Problems in the development of a test of functional literacy. Journal of Reading, 1977, 20(8), 697-705.

A bibliographic guide to functional literacy. Portland, OR: Northwest Regional Educational Laboratory, Clearinghouse for Applied Performance Testing, November 1979. (ERIC No. ED 189 197)

Bormuth, J. R. Reading literacy: Its definition and assessment. Reading Research Quarterly, 1974, 9(1), 7-66.

Bunch, M. B. Making decisions about adult learners based on performance on functional competency measures. Paper presented at the Adult Education Research Conference, San Antonio, 1978. (ERIC No. ED 154 137)

Caughran, A. M., & Lindlof, J. A. Should the "Survival Literacy Study" survive? Journal of Reading, 1972, 15(6), 429-435.

Cervero, R. M. Does the Texas Adult Performance Level Test measure functional competence? Adult Education, 1980, 30(3), 152-165.

Cervero, R. M. Assessment in adult basic education: A comparison of the APL and GED Tests. Adult Education, 1981, 31(2), 67-84.

Cronbach, L. J., & Meehl, P. E. Construct validity in psychological tests. Psychological Bulletin, 1955, 52, 281-302.

Final report: The Adult Performance Level Study. Austin, TX: University of Texas, Adult Performance Level Project, August 1977. (ERIC No. ED 185 113)

Fisher, D. L. Functional literacy tests: A model of question-answering and an analysis of errors. Reading Research Quarterly, 1981, 16(3), 418-448.

Fitzpatrick, A. R. The validity of criterion-referenced tests: Concepts and methods (Doctoral dissertation, University of Massachusetts, 1981). Dissertation Abstracts International, 1981, 42(5), 1091A. (Order No. 8117992)

Functional literacy: Basic reading performance. Denver, CO: National Assessment of Educational Progress, 1975.

Griffith, W. S., & Cervero, R. M. The Adult Performance Level Program: A serious and deliberate examination. Adult Education, 1977, 27(4), 209-224.

Haney, W. Validity and competency tests: The Debra P. case, conceptions of validity and strategies for the future. In G. F. Madaus (Ed.), The courts, validity, and minimum competency testing. Boston: Kluwer-Nijhoff, 1983.

Haney, W., & Lloyd, D. The APL Study: Science, dissemination, and the nature of adult education. In J. K. Fischer, W. Haney, & D. Lloyd, APL revisited: Its use and adaptation in states. Washington, D.C.: National Institute of Education, September 1980.

Harnisch, D. L. Pattern of item responses: An investigation with a functional literacy test. Paper presented at the Adult Education Research Conference, Lincoln, Nebraska, April, 1982.

Heller, B. R. New York State adult functional literacy models: Final report. Washington, D.C.: Office of Education (DHEW), 1978. (ERIC No. ED 145 100)

Hunter, C., & Harman, D. Adult illiteracy in the United States. New York: McGraw-Hill, 1979.

Kirsch, I., & Guthrie, J. T. The concept and measurement of functional literacy. Reading Research Quarterly, 1977-78, 13(4), 485-507.

Kronick, L. APL study. The Testing Digest, 1981, 3(2), 8-9, 15.

Lichtman, M. The development and assessment of R/EAL, and instrument to assess functional literacy. Journal of Reading Behavior, 1974, 2, 168-181.

Linn, R. L. Issues of validity in measurement for competency-based programs. In M. A. Bunda & J. R. Sanders (Eds.), Problems and practices in competency-based measurement. Washington, D.C.: National Council on Measurement in Education, 1979.

Madaus, G. F. Minimum competency testing for certification: The evolution and evaluation of test validity. In G. F. Madaus (Ed.), The courts, validity, and minimum competency testing. Boston: Kluwer-Nijhoff, 1983.

McClung, M. S. Competency testing: Potential for discrimination. Clearinghouse Review (National Clearinghouse for Legal Services), 1977, 11(5), 439-448.

Murphy, R. T. Adult Functional Reading Study: Final report. Washington, D.C.: National Institute of Education, 1973. (ERIC No. ED 109 650)

Nafziger, D. H., Thompson, R. B., Hiscox, M. D., & Owen, T. R. Tests of functional adult literacy: An evaluation of currently available instruments. Portland, OR: Northwest Regional Educational Laboratory, June 1975. (ERIC No. ED 109 265)

The 1971 National Reading Difficulty Index: A study of functional reading ability in the U.S. for the National Reading Center. New York: Louis Harris & Associates, 1971. (ERIC No. ED 057 312)

Olson, D. R. Review of Toward a literate society, edited by J. B. Carroll & J. S. Chall. Proceedings of the National Academy of Education, 1975, 2, 109-198.

Peck, C., & Kling, M. Adult literacy in the seventies: Its definition and measurement. Journal of Reading, 1977, 20(8), 677-682.

Pullin, D. Debra P. v. Turlington: Judicial standards for assessing the validity of minimum competency tests. In G. F. Madaus (Ed.), The courts, validity, and minimum competency testing. Boston: Kluwer-Nijhoff, 1983.

Rogers, S. K. Analysis of issues surrounding competency testing through examination of state statutes, court cases and educational literature (Doctoral dissertation, Kansas State University, 1981). Dissertation Abstracts International, 1981, 42(5), 1887A-1888A. (Order No. 8122031)

Sticht, T. G., & McFann, H. H. Reading requirements for career entry. In D. M. Nielsen & H. F. Hjelm (Eds.), Reading and career education (Perspectives in Reading No. 19). Newark, DE: International Reading Association, 1975.

Sticht, T. G., Caylor, J. S., Kern, R. P., & Fox, L. C. Project REALISTIC: Determination of adult functional literacy skill levels. Reading Research Quarterly, 1972, 7, 424-465.

Survival Literacy Study. New York: Louis Harris & Associates, 1970. (ERIC No. 068 813)

Willingham, W. W. New methods and directions in achievement measurement. New Directions for Testing and Measurement, 1980, 5, 73-80.